BIE-PST – Probability and Statistics

Lecture 2: Conditional probability and independence

Winter semester 2023/2024

Lecturer:
Francesco Dolce

Department of Applied Mathematics
Faculty of Information Technology
Czech Technical University in Prague

Table of contents

2 Conditional probability and independence

2.1 Conditional probability

Many statements about chance take the form "if B occurs, then the probability of A is p" where B and A are events.

How does the probability change if we have *partial information* about the result of the experiment?

Example 2.1. When rolling a balanced die with no additional information, we know that $\mathrm{P}(4) = 1/6$.

If we know that an even number was rolled, then it is clear that $\mathrm{P}(4 \mid \text{even}) = 1/3$.

Consider the *uniform* distribution on a set Ω with a *finite* "size" (e.g., the number of elements, length, area, capacity, time, etc.).

The probability of an event A is then defined as the by ratio of "sizes" as

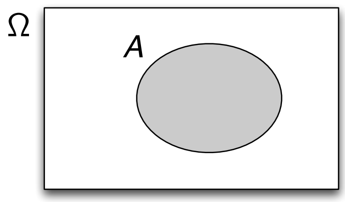
$$\mathrm{P}(A) = \mathrm{size}(A)/\mathrm{size}(\Omega).$$

If we know that an event B surely occurred, we are in fact interested only in outcomes of the experiment favorable to the event B. Favorable outcomes to the event A are now in $A \cap B$ and all of them must be in B (B surely occurred). We have

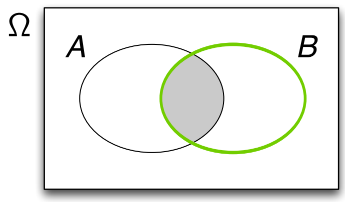
$$\mathrm{P}(A|B) = \frac{\mathrm{size}(A \cap B)}{\mathrm{size}(B)} = \frac{\mathrm{size}(A \cap B)/\mathrm{size}(\Omega)}{\mathrm{size}(B)/\mathrm{size}(\Omega)} = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}.$$

Definition 2.2. Let A, B be events and $\mathrm{P}(B) > 0$. The *conditional probability* of the event A given (the event) B is denoted by $\mathrm{P}(A|B)$ and is defined as

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}.$$



$$\mathrm{P}(A) = \frac{\mathrm{area}(A)}{\mathrm{area}(\Omega)}$$



$$\mathrm{P}(A \text{ given } B) = \frac{\mathrm{area}(\text{part of } A \text{ inside } B)}{\mathrm{area}(B)}$$

$$\mathrm{P}(A|B) = \frac{\mathrm{area}(A \cap B)\ /\ \mathrm{area}(\Omega)}{\mathrm{area}(B)\quad /\ \mathrm{area}(\Omega)}$$

$$\boxed{\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}}$$

$$\mathrm{P}(A \cap B) = \mathrm{P}(A|B)\ \mathrm{P}(B)$$
$$\mathrm{P}(A \cap B) = \mathrm{P}(B|A)\ \mathrm{P}(A)$$

$$\mathrm{P}(\text{intersection}) = \mathrm{P}(\text{event} \mid \text{condition})\ \mathrm{P}(\text{condition})$$

Example 2.3 (– rolling two dice)**.** Consider two rolls of a die. What is $\mathrm{P}(\text{sum} > 6 \mid \text{first} = 3)$?

The answer is surely $1/2$, since the second rolled number must be $4, 5$, or 6.

Formally: $\Omega = \{1, 2, 3, 4, 5, 6\}^2$,
$\mathrm{P}(A) = |A|/36$ for each $A \subset \Omega$.
Let $B = \{(3, \omega_2) : 1 \leq \omega_2 \leq 6\}$, $A = \{(\omega_1, \omega_2) : \omega_1 + \omega_2 > 6\}$.
Then

$$\mathrm{P}(A \mid B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \frac{\frac{|A \cap B|}{36}}{\frac{|B|}{36}} = \frac{|A \cap B|}{|B|} = \frac{3}{6}.$$

Example 2.4 (– family with two children)**.** A trickier example:
A family has two children. What is the probability that both are boys, given that at least one of them is a boy? I.e., what is the value of $\mathrm{P}(\text{both boys} \mid \text{at least one is a boy})$?

$\Omega = \{GG, GB, BG, BB\}$.

$$\begin{aligned}\mathrm{P}(BB|BG \cup GB \cup BB) &= \frac{\mathrm{P}(BB \cap (BG \cup GB \cup BB))}{\mathrm{P}(BG \cup BG \cup BB)} \\ &= \frac{\mathrm{P}(BB)}{\mathrm{P}(BG \cup GB \cup BB)} = \frac{1/4}{3/4} = \frac{1}{3}.\end{aligned}$$

<u>Incorrect:</u> $\mathrm{P}(BB|\text{older is boy}) = \mathrm{P}(BB|BG \cup BB) = \dfrac{\mathrm{P}(BB \cap (BG \cup BB))}{\mathrm{P}(BG \cup BB)} = \dfrac{1}{2}.$

Lemma 2.5. *Let* $\mathrm{P}(B) > 0$*. Then the conditional probability* $\mathrm{P}(\cdot|B)$ *is a probability measure, i.e.,* $\mathrm{P}(\cdot|B) \in [0, 1]$ *and it fulfills the axioms of probability.*

Proof. We need to prove the following:

i) $\mathrm{P}(\cdot|B) : \mathcal{F} \to \mathbb{R}$,

ii) *non-negativity:* for all $A \in \mathcal{F}$ it holds $\mathrm{P}(A|B) \geq 0$,

iii) *normalization:* $\mathrm{P}(\Omega|B) = 1$, $\quad \mathrm{P}(\Omega|B) = \dfrac{\mathrm{P}(\Omega \cap B)}{\mathrm{P}(B)} = \dfrac{\mathrm{P}(B)}{\mathrm{P}(B)} = 1,$

iv) $\sigma-$*additivity*: If $A_1, A_2, \ldots \in \mathcal{F}$ are mutually disjoint events (i.e., $A_i \cap A_j = \emptyset$ for $\forall i, j : i \neq j$), then

$$\mathrm{P}\left(\bigcup_{i=1}^{+\infty} A_i \middle| B\right) = \frac{\mathrm{P}\left(\left(\bigcup_{i=1}^{+\infty} A_i\right) \cap B\right)}{\mathrm{P}(B)} = \frac{\mathrm{P}\left(\bigcup_{i=1}^{+\infty} (A_i \cap B)\right)}{\mathrm{P}(B)} = \cdots = \sum_{i=1}^{+\infty} \mathrm{P}(A_i|B).$$

□

Conditional probability fulfills all mentioned properties of probability as well:

- if A_1 and A_2 are mutually disjoint, then $\mathrm{P}(A_1 \cup A_2|B) = \mathrm{P}(A_1|B) + \mathrm{P}(A_2|B)$,
- $\mathrm{P}(A_1 \cup A_2|B) = \mathrm{P}(A_1|B) + \mathrm{P}(A_2|B) - \mathrm{P}(A_1 \cap A_2|B)$,

- $\mathrm{P}(A^c|B) = 1 - \mathrm{P}(A|B)$,
- etc.

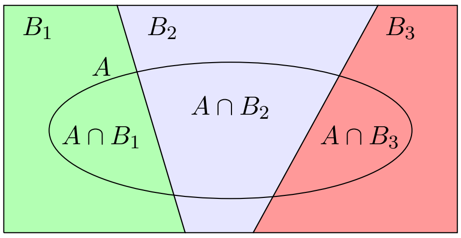
Moreover, the probability $\mathrm{P}(A|B)$ "lives" on B: for $A \cap B = \emptyset$ we have $\mathrm{P}(A|B) = 0$.

Furthermore, $\mathrm{P}(A \cap B|B) = \dfrac{\mathrm{P}(A \cap B \cap B)}{\mathrm{P}(B)} = \dfrac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \mathrm{P}(A|B)$.

2.2 Case distinct formula and Bayes' Theorem

Case distinct formula (Law of total probability)

$\Omega = B_1 \cup B_2 \cup B_3$ (disjoint partition)



Recall:

$$\boxed{\mathrm{P}(A|B_i) = \frac{\mathrm{P}(A \cap B_i)}{\mathrm{P}(B_i)}}$$

$$\boxed{\mathrm{P}(A \cap B_i) = \mathrm{P}(A|B_i)\,\mathrm{P}(B_i)}$$

$$A = A \cap \Omega = A \cap (B_1 \cup B_2 \cup B_3)$$

$$= (A \cap B_1) \cup (A \cap B_2) \cup (A \cap B_3)$$

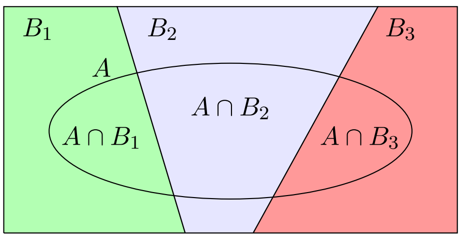
$$\mathrm{P}(A) = \ \mathrm{P}(A \cap B_1) \ + \ \mathrm{P}(A \cap B_2) \ + \ \mathrm{P}(A \cap B_3)$$

$$\boxed{\mathrm{P}(A) = \mathrm{P}(A|B_1)\,\mathrm{P}(B_1) + \mathrm{P}(A|B_2)\,\mathrm{P}(B_2) + \mathrm{P}(A|B_3)\,\mathrm{P}(B_3)}$$

Bayes' Theorem = converse procedure

At the end we observe A and we ask ourselves, what is the probability that the event B_j occurred.

$\Omega = B_1 \cup B_2 \cup B_3$ (disjoint partition)



Recall:

$$\boxed{\mathrm{P}(A \cap B_j) = \mathrm{P}(A|B_j)\,\mathrm{P}(B_j)} \quad \boxed{\mathrm{P}(A) = \mathrm{P}(A|B_1)\,\mathrm{P}(B_1) + \mathrm{P}(A|B_2)\,\mathrm{P}(B_2) + \mathrm{P}(A|B_3)\,\mathrm{P}(B_3)}$$

$$\mathrm{P}(B_j|A) = \frac{\mathrm{P}(A \cap B_j)}{\mathrm{P}(A)}$$

$$\boxed{\mathrm{P}(B_j|A) = \frac{\mathrm{P}(A|B_j)\,\mathrm{P}(B_j)}{\mathrm{P}(A|B_1)\,\mathrm{P}(B_1) + \mathrm{P}(A|B_2)\,\mathrm{P}(B_2) + \mathrm{P}(A|B_3)\,\mathrm{P}(B_3)}}$$

Bayes' Theorem (Thomas Bayes, 1701–1761)

A family of mutual disjoint events $B_1, B_2, \dots B_n$ is called a *partition* of the set Ω, if

$$\Omega = \bigcup_{i=1}^{n} B_i.$$

Theorem 2.6 (– case distinct formula (Law of total probability))**.** *Let $B_1, B_2, \dots, B_n$ be a partition of Ω such that $\forall\, i : \mathrm{P}(B_i) > 0$.*

Then for each event A it holds that

$$\mathrm{P}(A) = \sum_{i=1}^{n} \mathrm{P}(A|B_i)\,\mathrm{P}(B_i).$$

Theorem 2.7 (– Bayes' Theorem)**.** *Let $B_1, B_2, \dots, B_n$ be a partition of Ω such that $\forall\, i$: $\mathrm{P}(B_i) > 0$ and let A be an event with $\mathrm{P}(A) > 0$. Then it holds that*

$$\mathrm{P}(B_j|A) = \frac{\mathrm{P}(A|B_j)\,\mathrm{P}(B_j)}{\sum_{i=1}^{n} \mathrm{P}(A|B_i)\,\mathrm{P}(B_i)}.$$

Bayes' Theorem – example

Example 2.8 (– spam filter)**.** From the analysis of our email account we find out that:

- 30% of all delivered messages is spam;
- in 70% of spam messages there is the word "copy";
- only in 10% of non-spam messages there is the word "copy".

Calculate the probability that a message containing the word "copy" is a spam,

S: set of spam messages, $S^c = \Omega \setminus S$: set of non-spam messages, C: set of messages containing word "copy", C^c: set of messages not containing the word "copy".

$$\mathrm{P}(S) = 0.3,\ \mathrm{P}(S^c) = 0.7, \qquad \mathrm{P}(C|S) = 0.7,\ \mathrm{P}(C|S^c) = 0.1$$

$$\mathrm{P}(S|C) = \frac{\mathrm{P}(C|S)\,\mathrm{P}(S)}{\mathrm{P}(C|S)\,\mathrm{P}(S) + \mathrm{P}(C|S^c)\,\mathrm{P}(S^c)} = \frac{0.7 \cdot 0.3}{0.7 \cdot 0.3 + 0.1 \cdot 0.7} = \frac{21}{28} = 0.75.$$

2.3 Probability trees

Probability trees

First let us recall a useful relation: From the definition of conditional probability it follows that:

$$P(A \cap B) = P(A|B)\,P(B).$$

For 3 events it similarly holds that:

$$P(A \cap B \cap C) = P(A)\,P(B|A)\,P(C|A \cap B),$$

which can be proven by using the definition of conditional probability on the right hand side:

$$\begin{aligned} P(A)\,P(B|A)\,P(C|A \cap B) &= P(A)\frac{P(B \cap A)}{P(A)}\frac{P(C \cap (A \cap B))}{P(A \cap B)} \\ &= P(A \cap B \cap C). \end{aligned}$$

Lemma 2.9 (– Multiplicative law)**.** *Let for events $A_1, \dots, A_n$ hold that $P(A_1 \cap \dots \cap A_n) > 0$. Then it holds that*

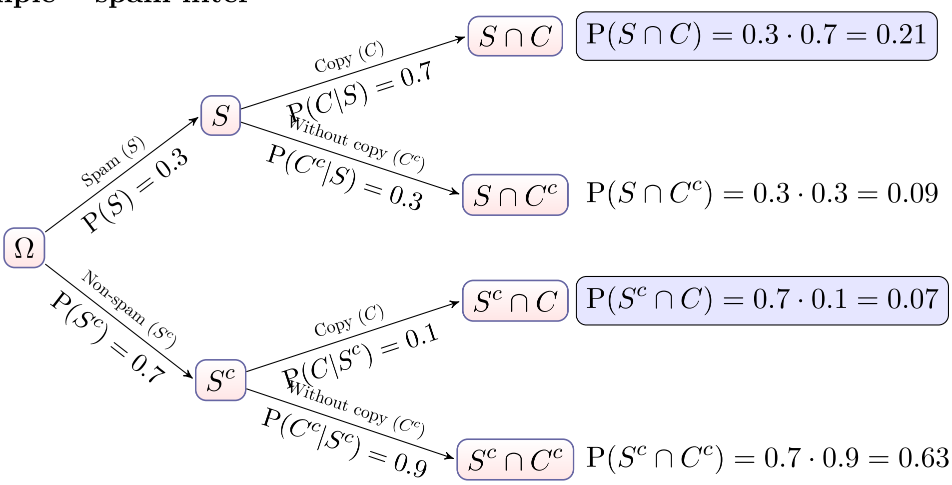
$$P(A_1 \cap \dots \cap A_n) = P(A_1)\,P(A_2|A_1)\,P(A_3|A_1 \cap A_2) \dots P(A_n|A_1 \cap \dots \cap A_{n-1}).$$

Proof. We apply successively the relation $P(A \cap B) = P(A)\,P(B|A)$ following from the definition of conditional probability:

$$\begin{aligned} P(A_1 \cap \dots \cap A_n) &= P(A_1 \cap \dots \cap A_{n-1})\,P(A_n|A_1 \cap \dots \cap A_{n-1}) \\ &= P(A_1 \cap \dots \cap A_{n-2})\,P(A_{n-1}|A_1 \cap \dots \cap A_{n-2})\,P(A_n|A_1 \cap \dots \cap A_{n-1}) \\ &= \dots . \end{aligned}$$

□

Example – spam filter

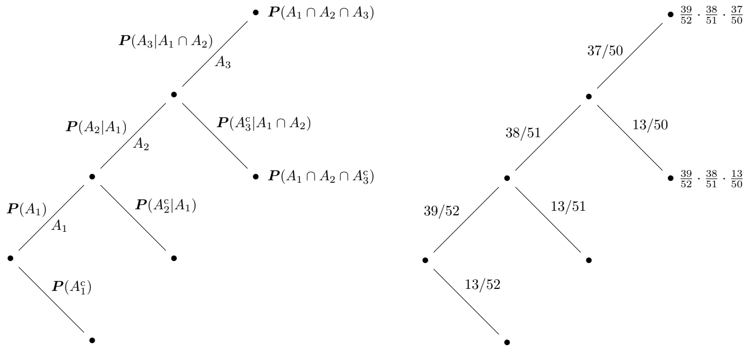


$$\mathrm{P}(S|C) = \frac{\boxed{\mathrm{P}(S \cap C)}}{\boxed{\mathrm{P}(C)}} = \frac{0.21}{0.21 + 0.07} = 0.75$$

Example 2.10 (– multiplicative law)**.** Suppose we draw cards without replacement from a 52 cards deck. What is the probability that in a sequence of 3 cards drawn one after another there are no hearts? $A_i = \{i\text{-th card is not hearts}\}, \quad i = 1, 2, 3.$

$\mathrm{P}(A_1 \cap A_2 \cap A_3) = \mathrm{P}(A_1)\,\mathrm{P}(A_2|A_1)\,\mathrm{P}(A_3|A_1 \cap A_2) = \frac{39}{52} \cdot \frac{38}{51} \cdot \frac{37}{50} \doteq 41.4\%.$

Illustration of computation by means of probability tree:



The probability of a given vertex of the tree is the product of the corresponding values on the path stemming from the root.

Misinterpretations of conditional probability

Many data misinterpretations and fallacies are based on incorrect understanding of conditional probabilities:

Example 2.11 (– driving under influence)**.**

- It was observed that approximatively 10% of fatal car accidents are caused by drunk drivers (46 out of 454 road fatalities in 2022 in the Czech Republic according to the yearly police report).
- This means that 90% of fatal accidents are caused by sober drivers!
- Does this mean that we should should beware of the sober drivers?

Of course not. We have to carefully read the probabilities.

The figure tells us that among all accidents, the percentage caused by drunk drivers is 10%. Thus

$$\mathrm{P}(\text{drunk}|\text{accident}) = 0.1.$$

What we are trying to find out is the reverse conditional probability $\mathrm{P}(\text{accident}|\text{drunk})$.

From a different study, we have found out that less than 1% of drivers are driving under influence. The overall chance of accident is difficult to determine, so we will compute just how more likely it is to cause an accident for drunk drivers:

$$\frac{\mathrm{P}(\text{accident}|\text{drunk})}{\mathrm{P}(\text{accident}|\text{sober})} = \frac{\mathrm{P}(\text{accident} \cap \text{drunk})/\,\mathrm{P}(\text{drunk})}{\mathrm{P}(\text{accident} \cap \text{sober})/\,\mathrm{P}(\text{sober})}$$
$$= \frac{\mathrm{P}(\text{drunk}|\text{accident}) \cdot \mathrm{P}(\text{accident})/\,\mathrm{P}(\text{drunk})}{\mathrm{P}(\text{sober}|\text{accident}) \cdot \mathrm{P}(\text{accident})/\,\mathrm{P}(\text{sober})} = \frac{0.1/0.01}{0.9/0.99} = 11.$$

Drunk drivers have at least 11 times higher probability of causing a fatal accident.

2.4 Independence of events

Independence of events

Intuitively: A and B are independent if the probability of the event A is not influenced by the knowledge about occurrence of the event B, i.e., $\mathrm{P}(A|B) = \mathrm{P}(A)$, and (vice versa) $\mathrm{P}(B|A) = \mathrm{P}(B)$.

Definition 2.12. Events A and B are called *independent*, if

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B).$$

Generally, a family of events $\{A_i \,|\, i \in I\}$ is called *independent* if

$$\mathrm{P}\left(\bigcap_{i\in J} A_i\right) = \prod_{i\in J} \mathrm{P}(A_i)$$

for all finite non-empty subsets J of I.

-

Example 2.13 (– rolling a die)**.** Consider the events A: "an even number is rolled" and B: "a number less than 3 is rolled".

Are the events A and B independent?

$$A = \{2,4,6\}, \qquad B = \{1,2\}, \qquad A \cap B = \{2\}.$$

$$\mathrm{P}(A \cap B) = \frac{1}{6} \qquad \text{and} \qquad \mathrm{P}(A)\,\mathrm{P}(B) = \frac{3}{6} \cdot \frac{2}{6} = \frac{1}{6}.$$

Then the events A and B *are independent.*

Example 2.14 (– rolling a die)**.** Consider the events A: "an even number is rolled" and B: "number 4 is rolled".

Are the events A and B independent?

$$A = \{2,4,6\}, \qquad B = \{4\}, \qquad A \cap B = \{4\}.$$

$$\mathrm{P}(A \cap B) = \frac{1}{6} \qquad \text{and} \qquad \mathrm{P}(A)\,\mathrm{P}(B) = \frac{3}{6} \cdot \frac{1}{6} = \frac{1}{12}.$$

Then events A and B *are <u>not</u> independent.*

Relation between independence and conditional probability

Let A and B be independent events and $\mathrm{P}(B) > 0$. Then clearly

$$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} = \frac{\mathrm{P}(A)\,\mathrm{P}(B)}{\mathrm{P}(B)} = \mathrm{P}(A).$$

For A and B independent the knowledge of B does not bring us any information about A.

Theorem 2.15. *If the events A and B are independent then A and B^c (resp., A^c and B; A^c and B^c) are independent, too.*

Theorem 2.16. *If $(A_i)_{i \in I}$ is a family of independent events, then for any arbitrary non-empty finite subset $\emptyset \neq J \subset I$ it holds that*

$$\mathrm{P}\left(\bigcap_{i \in J} A_i \mid \bigcap_{i \in I \setminus J} A_i\right) = \mathrm{P}\left(\bigcap_{i \in J} A_i\right).$$

Independent vs disjoint events

A common error is to make the fallacious statement that A and B are independent if $A \cap B = \emptyset$.

In fact, disjoint events A and B are independent only if $\mathrm{P}(A) = 0$ or $\mathrm{P}(B) = 0$.

If A and B are disjoint with non-zero probabilities, then the knowledge that B occurred tells us that A cannot occur.

The events being disjoint is a matter of sets, independence is a matter of probabilities.

Conditional independence

Definition 2.17. Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and C an event with $\mathrm{P}(C) > 0$. Events A and B are called *conditionally independent with respect to C*, if

$$\mathrm{P}(A \cap B|C) = \mathrm{P}(A|C)\,\mathrm{P}(B|C).$$

Recall:

- $Q(A) = \mathrm{P}(A|C)$ is a probability measure;
- the conditional independence is thus the independence with respect to probability Q.

Example 2.18 (– rolling a seven-sided die)**.** Suppose we roll a seven-sided die with all sides equally likely. Consider the events: A: "an even number is rolled", B: "a number less than 3 is rolled".

Are the events A and B independent? $A = \{2, 4, 6\}$, $B = \{1, 2\}, A \cap B = \{2\}$.

$$\mathrm{P}(A \cap B) = \frac{1}{7} \qquad \text{and} \qquad \mathrm{P}(A) \cdot \mathrm{P}(B) = \frac{3}{7} \cdot \frac{2}{7} = \frac{6}{49}.$$

Events A and B *are not independent.*

Example 2.19 (– rolling a seven-sided die + condition)**.** Consider further event C: "we rolled at most 6" $C = \{1, 2, 3, 4, 5, 6\}$.

Are events A and B conditionally independent with respect to C?

$$\mathrm{P}(A \cap B|C) = \frac{\mathrm{P}(A \cap B \cap C)}{\mathrm{P}(C)} = \frac{\mathrm{P}(\{2\})}{\mathrm{P}(\{1, \ldots, 6\})} = \frac{1/7}{6/7} = \frac{1}{6},$$

$$\mathrm{P}(A|C) \cdot \mathrm{P}(B|C) = \frac{3/7}{6/7} \cdot \frac{2/7}{6/7} = \frac{1}{6}.$$

Events A and B *are conditionally independent* with respect to C.

